

Tundra Swan State Space Model

Erik Osnas

25 April, 2017

Below, I describe a Bayesian state-space model on which to base swan population assessment and as an alternative to a lagging three-year average of a population index. I also describe basic properties of Bayesian state-space models and how the model “smooths” the underlying state variables. While the three-year average is a straightforward method to smooth a variable index that contains sampling error, the route use of three years is essentially arbitrary. Why not a 2 or 5 year average? A 2-year average is less smooth and a 5-year average is more smooth. The appropriate amount of smoothing should depend on our understanding of the measure error in the index, the true fluctuations in the population, and our biological knowledge of the underlying population. A Bayesian state-space model is a method to combine these to predict or estimate the underlying population state.

Plot Swan Data

Below is a plot of the swan index data.

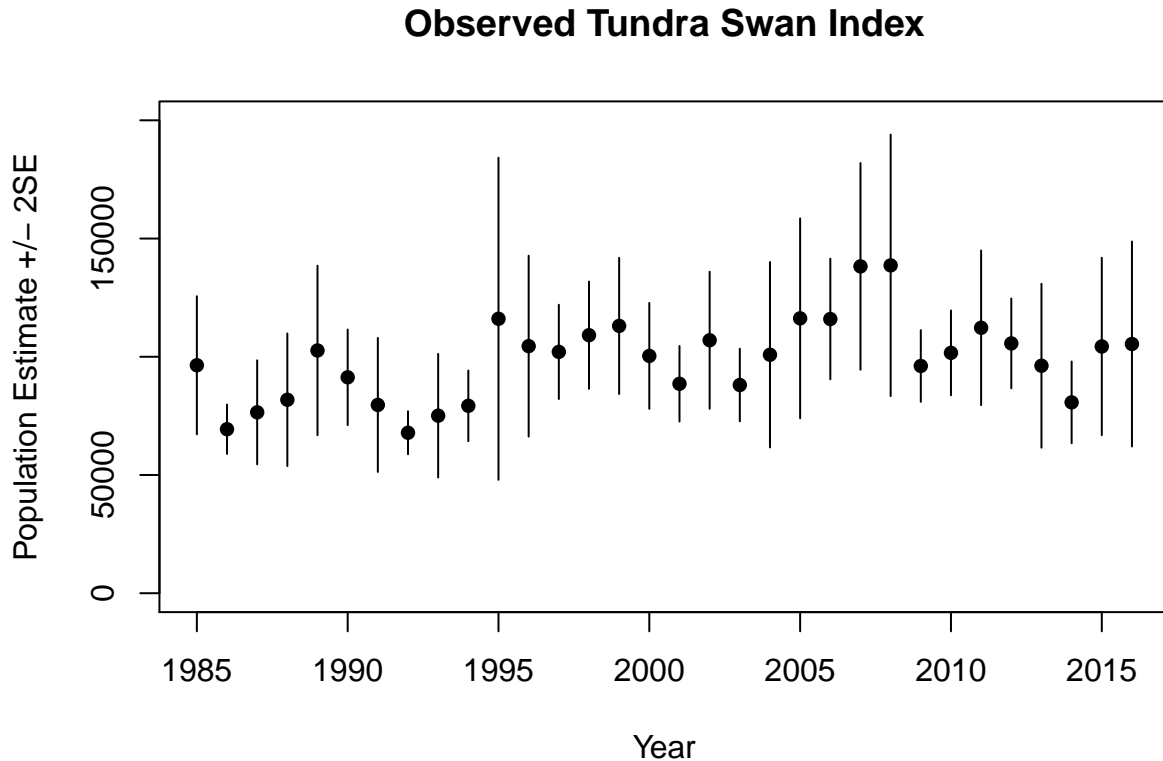


Figure 1: Swan population index from 1985 to 2016.

Note that some years have a high estimate followed by years with low estimates or visa verse. Many of the high estimates also have high error estimates. This variation is a product of sampling (measurement) error

and true underlying changes in the observed population from year-to-year (process error) in the sampled region. Because the standard error estimates give the sampling error, we can use these in combination with a population model to estimate the process error and the underlying population state.

Fitting a Bayesian State Space Model

Here we fit a Bayesian state-space model to the swan data. A state-space model is a general term that usually refers to a model containing the true underlying (and unobserved) time-dependent state of the system and imperfect time-dependent observations. The states change with time so that the state at time t depends on the state at time $t - 1$. Here, observations depend just on the state, but more complicated observation processes could be modeled. State-space models were first developed as part of missile guidance systems and the Apollo space program in the 1950s and 60s (the Kalman filter, see en.wikipedia.org/wiki/Kalman_filter), so in this sense a state-space model can be thought of as “rocket science”. Ecological processes, however, are often more difficult to predict than rocket trajectories.

The model is Bayesian only because we implement it as such, but this also allows us to use prior information in the estimation of model parameters. You can think of prior information as our biological knowledge that is more general than any specific data set. You can also think of priors as a set of assumptions we make in order to use the methodology. When priors are based on highly specific and certain knowledge, they are called “informative” and can be very influential to the analysis.

The model has four parts: a mathematical population projection model, and observation model, priors on all parameters, and observed data. The population projection model is

$$N_{t+1} = N_t e^{r_t}$$

with

$$r_t \sim \text{Normal}(\bar{r}, \sigma_1).$$

The \sim means “is distributed as” and signifies a random quantity. Thus, the underlying model is a stochastic exponential growth model, as σ_1 goes to zero, the population trajectory goes to the curve defined by $N_{t+1} = N_t e^{r_t}$; as σ_1 increases, the trajectory departs more from a fixed curve and looks more and more random as σ_1 increases further. These departures from the modeled curve are not “error” in the sense of random unrepeatable effects or measurement error, but represent true underlying departures from the mathematical model; therefore this component of the variance is often called “process variance” or “process error” in the sense that these departure are “error” with respect to the mathematical model but still represent true population fluctuation from all causes not modeled. In other words, these deviations represent departures from the deterministic model.

The observation component of the model is

$$Y_t \sim \text{Normal}(N_t, \sigma_{2,t})$$

Here Y_t are the observed data (index values), and the standard deviation, $\sigma_{2,t}$, is indexed by time, meaning that it is year-specific. Because this is also calculated as part of the population index, $\sigma_{2,t}$ is input as data along with the index observations. The difference between N_t and Y_t represent errors in the sense of sampling noise—if the index survey was repeated with a different sample of transects and calculated anew (but all else being the same), the observed index value would be different. This error reflect our inability to observe the system perfectly. Having an estimate of the sampling variance (observation error) allows us to separate process variance from sample variance and to “smooth” the population estimate based on the data.

The final component of the model is a set of priors for the parameters. Priors represent our belief before we analyze the current data. One possible set of priors for the parameters is

$$\bar{r} \sim \text{Normal}(0, 0.1),$$

$$\sigma_1 \sim \text{Uniform}(0, 0.3), \text{ and}$$

$$\log(N_1) \sim \text{Normal}(\log(75000), 0.1).$$

The parameter \bar{r} represents the long term average growth rate of the population on the log scale. The prior is normally distributed and centered on zero but could vary such that there is an approximately 95% probability that the true growth rate is ± 0.2 units from zero each year. At these extremes, the population is changing about $\pm 20\%$ per year over more than 30 years. An average growth rate that is near $\pm 20\%$ would mean the population is quickly going extinct or will outgrow the available space on planet Earth. Clearly the true long term average growth rate is nearer to zero than these extremes, and this belief is reflected in the choice of a normal distribution with a mean of zero. The prior for parameter σ_1 is Uniform (equally probable) between 0 and 0.3 on the log population scale (standard deviation parameters can only be positive). This equates approximately to an annual coefficient of variation of about 30% at the maximum; meaning that if σ_1 is 0.3, then on average the annual deviation between the unobserved true population and the mathematical model will be about 30%. Any one deviation in a given year will be very different because it is a random quantity with a normal distribution on the log scale. On average across a large number of observations, however, deviations will be about 30% different from the mathematical model if σ_1 is set at the maximum. Finally, we need to specify a prior for the initial population size in the first year of observation. This basically constrains the model near biologically realistic populations sizes. Without this constraint, models of this type can be very difficult to fit. Here, the initial (log) population size is given a normal prior centered on 75000 swans and a coefficient of variation of about 10%. There is usually very little information to inform (estimate) this value in actual data sets.

Final estimates from a Bayesian analysis depend on both the priors and the data. When little or no information is contained in the data, estimates will to a greater degree reflect the prior. Conversely, when the data contributes much information, the final estimates (called “posterior estimates”) can be very different from the priors. For time series-type data and when estimating variance components, as is the case here, 30 years of data can be thought of as mildly informative—not bad but also not great.

In the software language of BUGS (Lunn, et al. 2009) or JAGS (Plummer 2003) used through R (R Core Team 2015), the above model is represented as:

```
cat("model{
# Priors
  logN.est[1] ~ dnorm(log(75000), 1/(0.1)^2) # Prior for initial population size log scale
  mean.r ~ dnorm(0, 1/(0.1)^2)              # Prior for mean growth rate
  sigma.proc ~ dunif(0, 0.3)                # Prior for sd of state process log scale
  tau.proc <- pow(sigma.proc, -2)
# Likelihood
# State process
for (t in 1:(T-1) ) {
  r[t] ~ dnorm(mean.r, tau.proc)
  logN.est[t+1] <- logN.est[t] + r[t]
}
# Observation process
for (i in 1:T) {
  tau.obs[i] <- pow(sigma.obs[i], -2)
  y[i] ~ dnorm(exp(logN.est[i]), tau.obs[i])
}
}",
  file = "ssm.jags", fill = TRUE)
```

The above code writes a file to the directory, which is then used by BUGS or JAGS to define a simulation model that provides Bayesian posterior estimates. In R, we need to set up a list of data, a function that supplies initial values for each simulation, and then call and run the model. We run 3 independent simulations (“chains”) that allow us to assess model convergence and be sure our estimates do not depend on the initial values supplied. In R using the package R2jags:

```

# structure data
jags.data <- list(T = length(dat$Year), y = dat$Bpop,
  sigma.obs = dat$Sebpop)
# Initial values
inits <- function() {
  list(logN.est = c(runif(1, log(74000), log(76000)),
    rep(NA, 31)), mean.r = runif(1, -1e-04, 1e-04),
    sigma.proc = runif(1, 0.01, 0.011), r = c(runif(31,
      -0.01, 0.01)))
}
# Parameters monitored
parameters <- c("logN.est", "mean.r", "sigma.proc")
# Call JAGS from R
out <- jags(jags.data, inits, parameters, model.file = "ssm.jags",
  n.chains = 3, n.thin = 1, n.iter = 11000, n.burnin = 10000,
  working.directory = getwd())

```

```

## module glm loaded

## Compiling model graph
##   Resolving undeclared variables
##   Allocating nodes
## Graph information:
##   Observed stochastic nodes: 32
##   Unobserved stochastic nodes: 34
##   Total graph size: 243
##
## Initializing model

```

A plot of some output and diagnostic information is examined to make sure the model worked (Figure 2). Inspecting the “R-hat” diagnostic, we see that it is near 1.0, which means that each chain is fluctuating around a similar distribution (the posterior).

Now that we are confident that the model simulations settle on a posterior, we can plot the data and estimated parameters. The observed index values (data) with the posterior mean of the estimated population size (N_t) and the three-years average of the observed data are shown in Figure 3.

We can see that the posterior mean of the modeled population is less variable than the three-year average (Figure 3, Table 1). This will not be the case for all data sets. In this model, the population estimates are more smooth than the three-year average because the years with high point estimates for the index also have large standard errors (Figure 3). This causes the model to weight these data points less than others when estimating the modeled population. The degree of weight is determined by the observation variance that is input as data and on the estimated process variance. Because these parameters are estimated from the full time series of data and the priors, which themselves are based on our biological knowledge, the degree of smoothing is “based on all the data”, explicit estimates of biological parameters (growth rate and its variance), and an explicit statement of our biological assumptions (the mathematical model and priors).

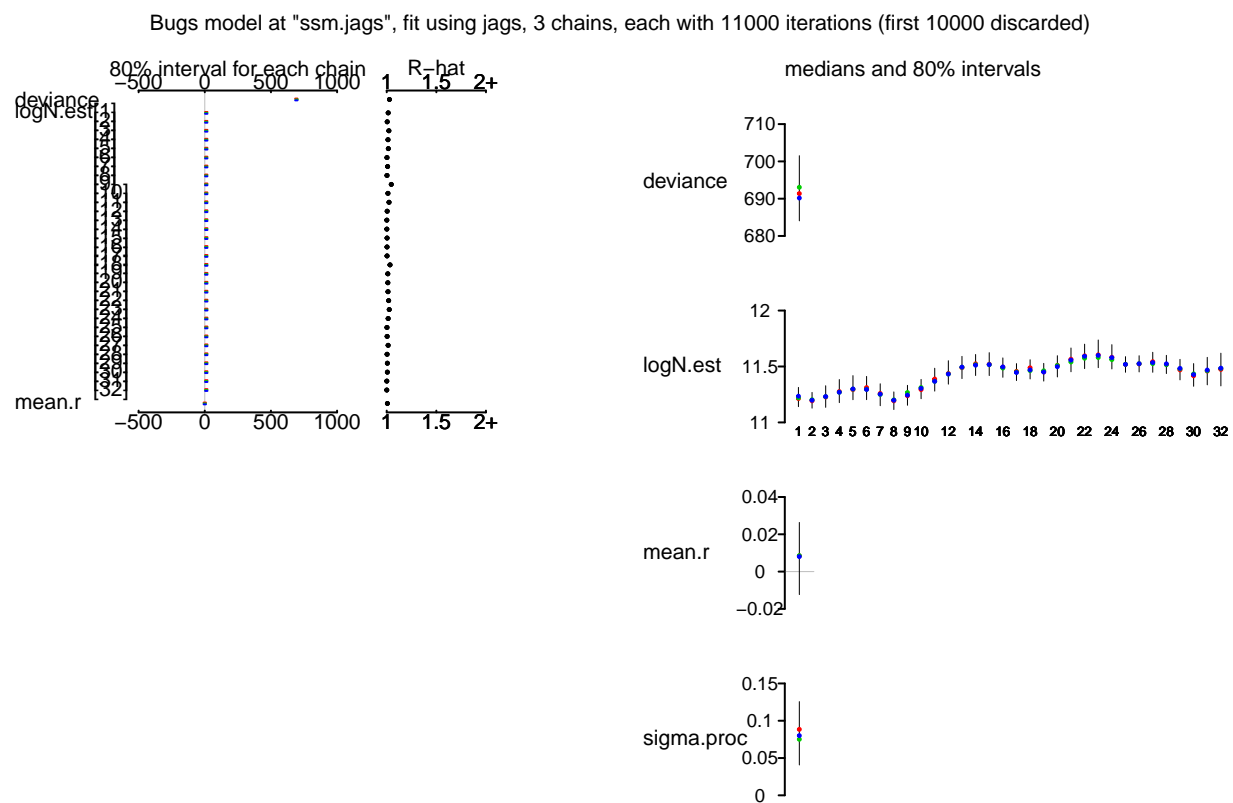


Figure 2: Model diagnostic output to check convergence of the posterior simulations.

Observed Tundra Swan Index and Bayesian Estimates

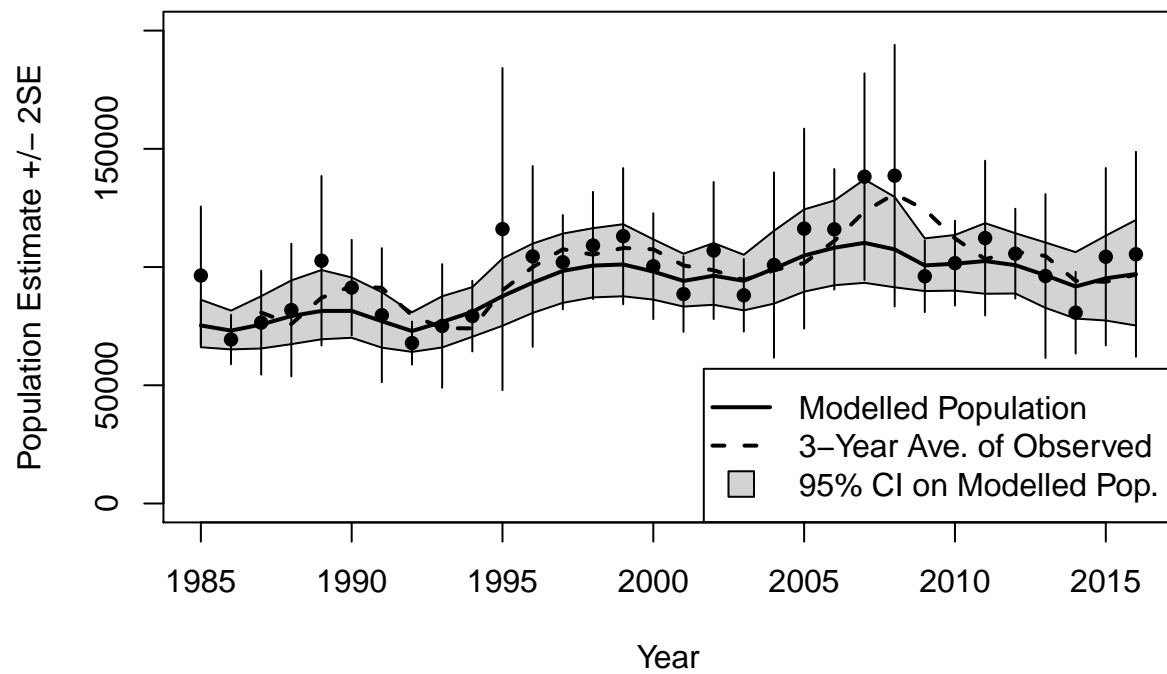


Figure 3: The observed swan index data plotted with the estimated posterior mean and 95% credible interval for the modelled population state and trailing three-year average of the observed index.

Table 1: The observed swan index (Observed Index), the standard deviation calculated from the indec (SD Index), the posterior mean estimated population size from the Bayesian model described in the text (Posterior Mean Population), the standard deviation of the population size estimate (Posterior SD Population), and the trailing three year average of the observed index (Three Year Average).

Year	Observed Index	SD Index	Posterior Mean Population	Posterior SD Population	Three Year Average
1985	96409	14599	75303	5147	NA
1986	69363	5228	73120	4045	NA
1987	76503	11002	75686	5618	80758
1988	81816	14027	79292	6756	75894
1989	102673	17923	81441	7178	86997
1990	91317	10106	81454	6603	91935
1991	79634	14177	77095	5902	91208
1992	67855	4547	72932	4409	79602
1993	75084	13069	76816	5538	74191
1994	79257	7470	81016	5352	74065
1995	116070	34057	87701	7374	90137
1996	104493	19122	93394	7683	99940
1997	102061	9970	98382	7545	107541
1998	109124	11341	100624	7416	105226
1999	113055	14401	101094	7987	108080
2000	100364	11195	98093	6584	107514
2001	88578	7987	94120	5586	100666
2002	106993	14491	96335	6650	98645
2003	88054	7675	94195	5854	94542
2004	100853	19603	99306	7768	98633
2005	116252	21140	104857	8947	101720
2006	115958	12750	108181	9185	111021
2007	138227	21856	110224	11199	123479
2008	138650	27658	107483	9824	130945
2009	96108	7584	100692	5662	124328
2010	101655	8972	101410	5839	112138
2011	112259	16360	102535	7418	103341
2012	105648	9501	100799	6461	106521
2013	96203	17331	96527	6953	104703
2014	80706	8649	91767	7184	94186
2015	104359	18757	95222	9100	93756
2016	105405	21676	97060	11316	96823

Posterior estimates of two important biological parameters are shown as histograms for mean growth rate and process variance (Figure 4). We can see that the posterior for both mean growth rate and the process standard deviation is very different than the priors for each parameter (Figure 4). Recall that the prior for mean growth rate was $\text{Normal}(0, 0.1)$ but that the posterior has a mean and standard deviation closer to 0.01. For the process variance, the prior was $\text{Uniform}(0, 0.3)$; whereas, the posterior mean is 0.08 and the 95% quantile is 0.14.

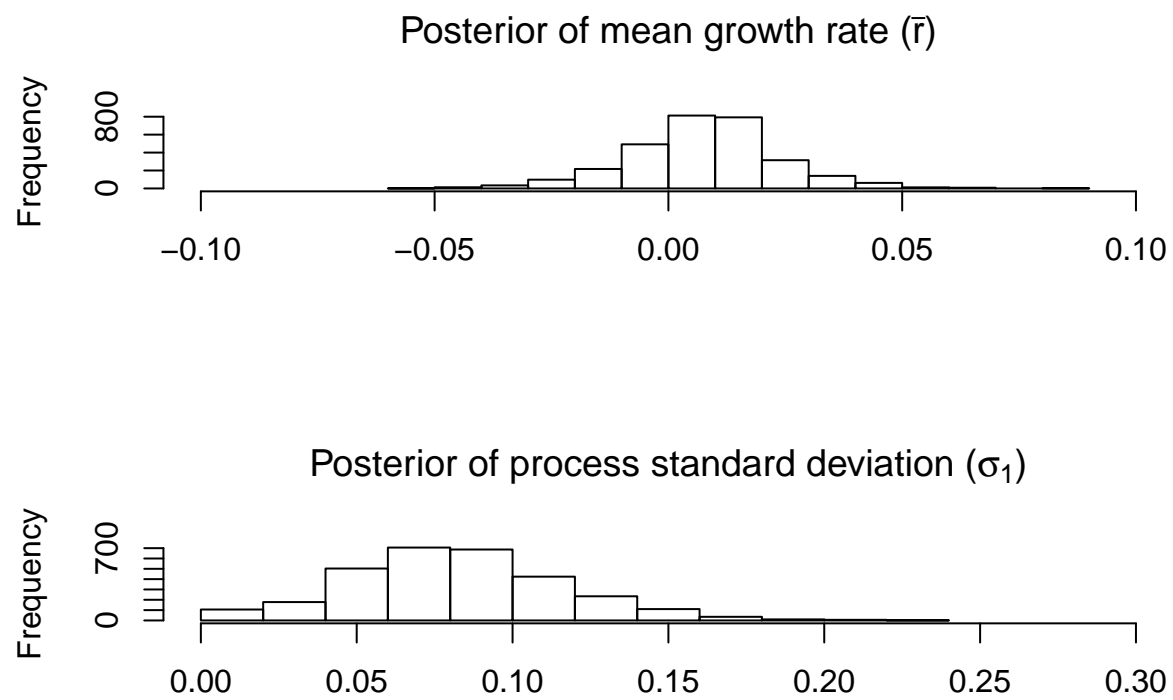


Figure 4: Posterior histograms for mean growth rate and process variance

Sensitivity Analysis on Priors

Here I vary the priors to see how large of an effect different prior might have. I only tried one different set that is much more informative for the mean growth rate and process variance and less informative for the initial population size. Before this model is adopted, additional thought should be given to priors.

```
cat("model{
# Priors
logN.est[1] ~ dnorm(log(75000), 1/(0.2)^2) # Prior for initial population size log scale
mean.r ~ dnorm(0, 1/(0.05)^2)             # Prior for mean growth rate
sigma.proc ~ dunif(0, 0.1)                 # Prior for sd of state process log scale
tau.proc <- pow(sigma.proc, -2)
# Likelihood
# State process
for (t in 1:(T-1)){
  r[t] ~ dnorm(mean.r, tau.proc)
  logN.est[t+1] <- logN.est[t] + r[t]
}
# Observation process
for (i in 1:T) {
  tau.obs[i] <- pow(sigma.obs[i], -2)
  y[i] ~ dnorm(exp(logN.est[i]), tau.obs[i])
}
}",
  file = "ssm.2.jags", fill = TRUE)
```

Convergence is checked (Figure 5), and then the data and estimated parameters with new priors are plotted (Figure 6). As can be seen (Figure 6), there is very little effect of these different priors on the plotted results, or the smoothing. The posterior histogram of mean growth rate is also similar to above, but process variance is truncated at 0.1 due to the prior (figure 7). Despite this difference, there was very little practical effect on the smoothing.

Discussion

Although Bayesian state-space models are slightly more complex than calculating a lagging average, they provide an empirical and theoretical justification for the amount of index smoothing. Therefore, management decision rules can be based directly off the smoothed population index at the posterior mean or other posterior quantity. Despite the increased complexity, it is becoming routine to fit such model since the development of software programs in the 1990s (e.g., BUGS, JAGS, and now STAN). The current simulations took less than a minute to run per model and can easily be automated to produce results in a timely fashion for annual reports. More complex models could also be developed but at the expense of computation time and additional parameters, model assumptions, and priors.

One advantage of Bayesian state-space models not discussed above is the ability to estimate missing data. In the event that a survey is not completed, then the state-space model can predict what observation would have been made and the underlying state of the population, just as for other parameters. These estimates are based on the posterior distributions of all other parameters; therefore, the estimates of missing data are a product of the non-missing data, mathematical model, and the priors just as for other parameters. In addition, these predictions of missing data also contain the same degree of smoothing as the rest of the model and fully account for all uncertainties in parameter estimation. In the same way as missing data, predictions of the future population state can be made any number of years in the future.

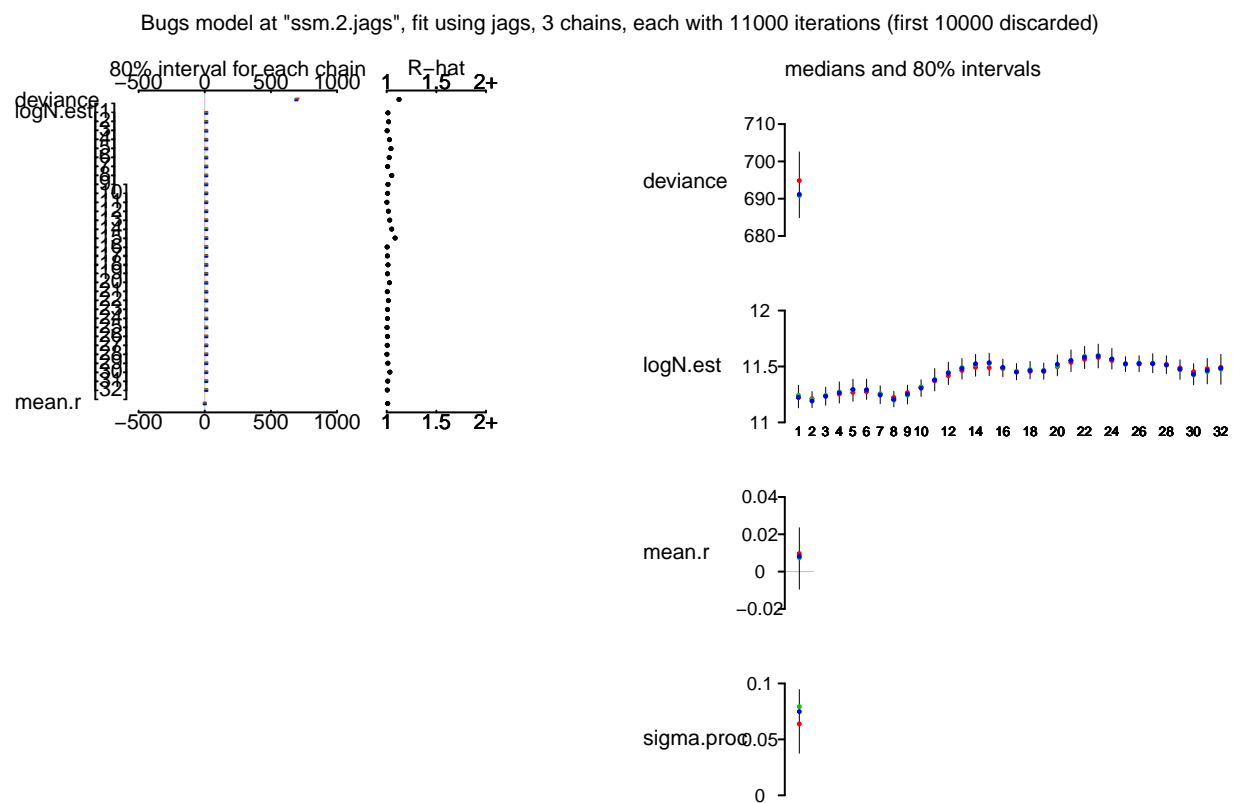


Figure 5: Model diagnostic output to check convergence of the posterior simulations.

Observed Tundra Swan Index and Bayesian Estimates

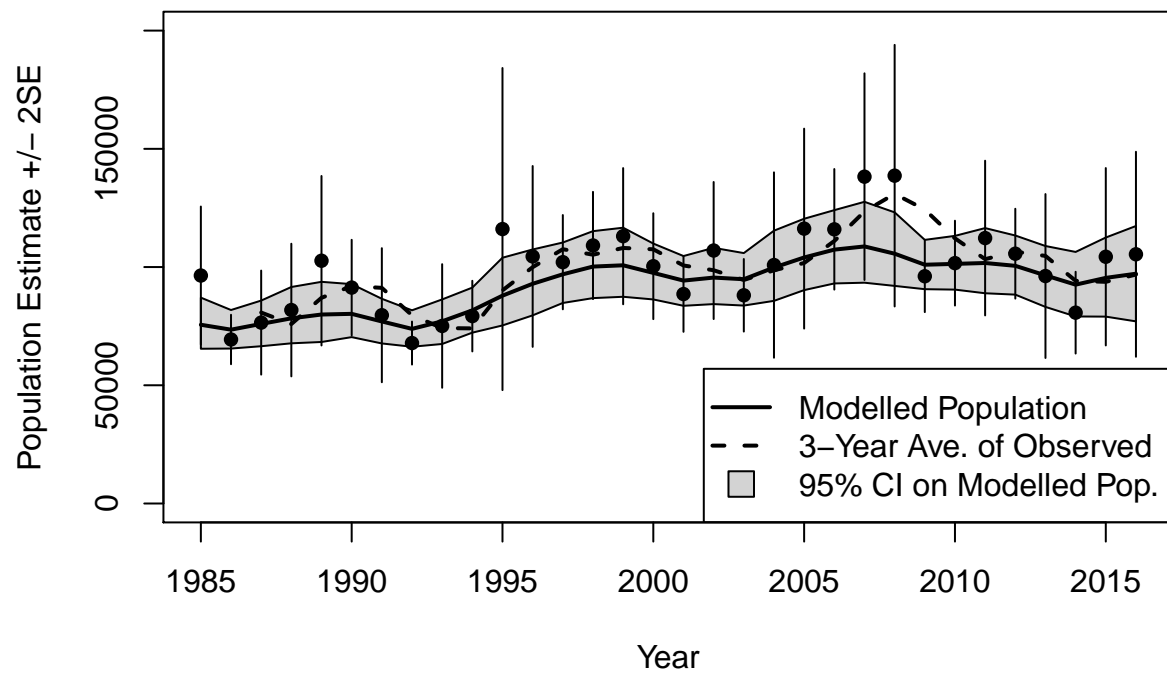


Figure 6: The observed swan index data plotted with the estimated posterior mean and 95% credible interval for the modelled population state and trailing three-year average of the observed index.

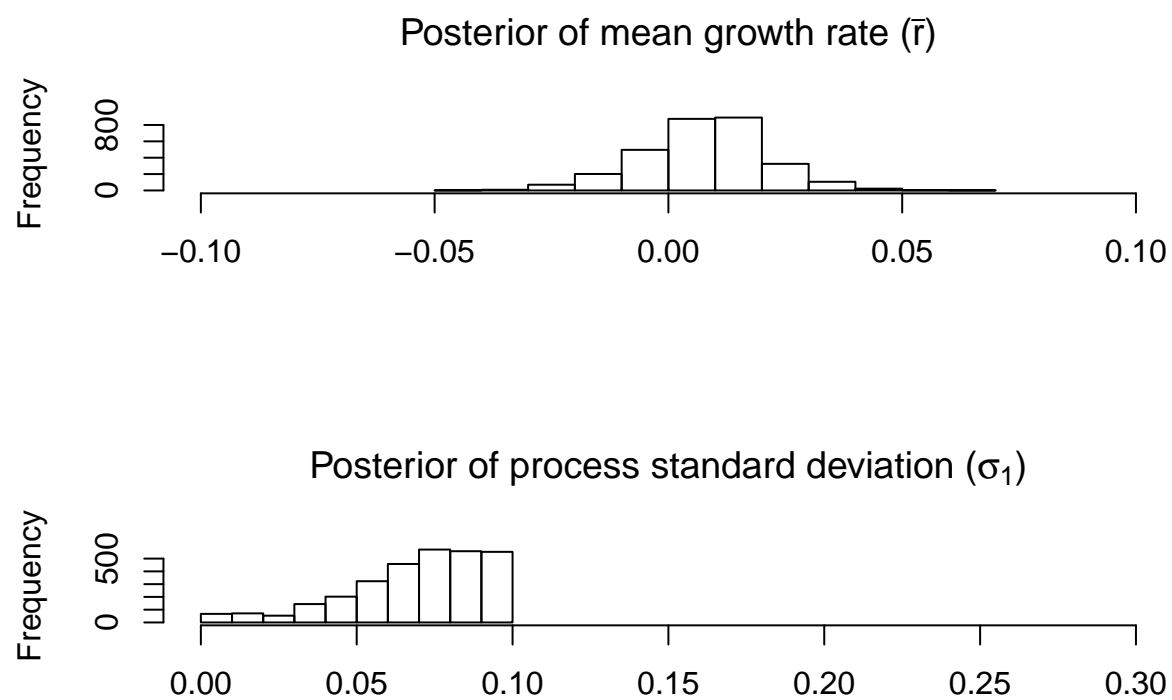


Figure 7: Posterior histograms for mean growth rate and process variance

References

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. (2009). “The BUGS project: Evolution, critique and future directions”. *Statistics in Medicine*. 28 (25): 3049-3067.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20-22, Vienna, Austria.